

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

<https://llrjournal.com/index.php/11>

Development of Lexical Bundles in Pakistani ESL Learners' Academic Writing across CEFR Levels: A Corpus-Based Study



Sadia Parveen

Lecturer Department of English Linguistics, IUB Bahawalpur
sadia.parveen@iub.edu.pk

Dr. Shahid Nawaz

Assistant Professor Department of English Linguistics, IUB Bahawalpur
shahidnawaz@iub.edu.pk

Abstract

This study investigates the development of lexical bundles in Pakistani ESL learners' academic writing across B1, B2, and C1 proficiency levels of the Common European Framework of Reference for Languages. Lexical bundles are recurrent multi-word sequences that support fluency, coherence, and rhetorical organization in academic discourse. Using a corpus-based approach, the study examines three-word and four-word lexical bundles in argumentative essays written by Pakistani ESL learners. The analysis focuses on bundle frequency, structural patterns, functional categories, and developmental movement across proficiency levels. The findings indicate that B1 learners rely mainly on basic and repetitive bundles, often connected with personal stance and simple textual organization. B2 learners show a transitional pattern, with greater use of discourse-organizing and explanatory bundles that help in argument development. At the C1 level, learners demonstrate more controlled and academic use of lexical bundles, including patterns associated with evidence, interpretation, and formal academic expression. The study shows that lexical bundle use develops gradually with language proficiency and reflects learners' growing ability to organize ideas in academically acceptable ways. The findings have direct relevance for ESL writing instruction in Pakistan, particularly for curriculum design, writing pedagogy, and corpus-informed teaching of academic phraseology.

Key words: lexical bundles, Pakistani ESL learners, academic writing, CEFR, corpus-based study, argumentative essays, academic phraseology

Introduction

Academic writing has a central place in higher education because it shows how learners organize ideas, develop arguments, and use language according to the expectations of formal academic discourse. For English as a Second Language learners, academic writing is often more demanding because they are required to manage grammar, vocabulary, coherence, argumentation, and academic style at the same time. In Pakistan, English carries special importance in universities, examinations, official communication, and professional mobility. However, many learners still face problems in producing fluent and well-organized academic texts. Their writing often depends on memorized expressions, limited vocabulary, repetitive sentence patterns, and classroom-based formulae. This situation makes it necessary to study the actual linguistic patterns used by Pakistani ESL learners, especially those patterns that support fluency, coherence, and academic organization.

One important area in this connection is lexical bundle research. Lexical bundles are recurrent sequences of words that appear frequently in a particular register, such as on the other hand, as a result of, it is important to, and the purpose of this. These bundles are not always complete grammatical units, but they perform important functions in discourse. They help writers link ideas, introduce claims, present evidence, mark contrast, and guide readers through an argument. Biber et al. (2016) described lexical bundles as frequent word combinations that act as building blocks of discourse. Hyland (2008) also argued that lexical bundles are important in academic writing because they help writers manage textual organization, reader engagement, and research-related

expression. In this way, lexical bundles are not simply repeated phrases; they are part of the phraseological system through which academic writing becomes more fluent and recognizable.

For ESL learners, lexical bundle use is closely linked with writing development. Learners at lower proficiency levels often rely on a narrow range of familiar bundles, while more advanced learners tend to use a wider and more contextually suitable range of academic bundles. This development is not only a matter of using more phrases. It also involves a shift in how learners use these phrases to perform academic functions. For example, a lower-level learner may repeatedly use *in my opinion* or *on the other hand* without developing a complex argument, while a more advanced learner may use bundles such as *the findings suggest that* or *it can be argued that* to express evidence-based reasoning. Hyland and Jiang (2018) noted that academic writers depend on recurring phraseological patterns to create coherence and disciplinary acceptability. Such patterns are especially useful for ESL learners because they reduce the burden of constructing every sentence from the beginning and allow learners to focus more on meaning and argument.

The Pakistani ESL context provides an important site for such research. Much of the existing work on lexical bundles has been conducted in native-speaker contexts or among learners from Western and East Asian backgrounds. Pakistani learners, however, write English within a multilingual setting shaped by Urdu, regional languages, examination traditions, and classroom practices. Their academic writing may show patterns that are different from those found in other ESL contexts. Since English is widely used as a medium of instruction in higher education, learners are expected to write academic essays, assignments, reports, and research papers in English. Yet many of them receive limited formal instruction in academic phraseology. This gap between institutional expectations and classroom preparation makes lexical bundle development a useful area of inquiry.

The Common European Framework of Reference for Languages provides a useful basis for examining this development across proficiency levels. The CEFR classifies language ability into levels, allowing researchers to compare learners' writing in a more systematic way. In the present study, the focus is on B1, B2, and C1 levels. These levels represent a movement from intermediate writing ability toward more advanced academic control. At B1 level, learners are expected to produce simple connected texts. At B2 level, they are expected to present clearer arguments and more detailed writing. At C1 level, learners are expected to express ideas with greater fluency, organization, and precision. When lexical bundles are examined across these levels, it becomes possible to see how learners' academic phraseology changes as their proficiency improves.

A corpus-based approach is suitable for this purpose because it allows the researcher to examine real language data rather than depending only on general impressions of learner writing. Through corpus analysis, recurrent three-word and four-word bundles can be identified, counted, classified, and compared across proficiency levels. This method also makes it possible to examine structural and functional patterns. Structurally, lexical bundles may appear as noun-based, verb-based, or preposition-based forms. Functionally, they may be research-oriented, text-oriented, or participant-oriented, following Hyland's (2008) classification. Such analysis can show not only which bundles are used frequently but also how they contribute to the organization and meaning of academic writing.

The present article investigates lexical bundle development in Pakistani ESL learners' academic writing across CEFR levels. It focuses on three main concerns: the frequent three-word and four-word lexical bundles used by learners at B1, B2, and C1 levels; the structural and functional differences in lexical bundle use across these levels; and the way lexical bundle development reflects growth in academic writing proficiency. By focusing on Pakistani ESL learners, the study contributes to corpus-based research from a South Asian context and offers useful directions for English language teaching, writing instruction, and curriculum planning in Pakistan.

Literature Review

Lexical bundles have become an important area of study in corpus linguistics because they show how academic writers use recurring multi-word sequences to organize meaning, connect ideas, and develop arguments. These bundles are usually identified as three-word or four-word combinations that occur frequently across a corpus. Although they may not always form complete grammatical units, they carry clear discourse functions in academic writing. Biber et al. (2016) described lexical bundles as repeated word sequences that work as building blocks of discourse, while more recent work has continued to show that these bundles help writers produce fluent and coherent academic texts (Biber et al., 2021; Hyland & Jiang, 2018). In academic writing, expressions such as *on the other hand*, *as a result of*, *in the case of*, and *it is important to* are not accidental combinations; they guide readers through claims, contrasts, explanations, and evidence.

The role of lexical bundles becomes more significant in ESL academic writing because second language learners often face difficulty in producing texts that are both linguistically accurate and rhetorically organized. Academic writing demands more than grammar and vocabulary; it also requires control over conventional patterns of expression. Hyland (2008) argued that lexical bundles help writers manage academic discourse by organizing text, presenting research, and engaging readers. Recent studies have also noted that learners' ability to use lexical bundles appropriately reflects their growing command of academic phraseology and discourse competence (Chen & Baker, 2016; Hyland & Jiang, 2018). For ESL learners, lexical bundles may reduce the burden of sentence construction and provide ready-made frames for argument development, but this benefit depends on whether learners understand the function of the bundles they use.

Research has shown that lexical bundle use varies across proficiency levels. Lower-level learners often depend on a narrow and repetitive set of bundles, many of which are closer to spoken or personal expression, such as *in my opinion*, *I think that*, and *on the other hand*. These bundles may help learners connect ideas, but repeated and mechanical use can make writing less formal and less persuasive. More proficient learners usually show a wider range of text-oriented and research-oriented bundles, such as *the results show that*, *the findings suggest that*, and *in relation to the*. Chen and Baker (2016) found that learner writing differs from more advanced academic writing in the range, structure, and function of lexical bundles. Hyland and Jiang (2018) also observed that phraseological development is shaped by exposure to academic texts, writing practice, and awareness of rhetorical conventions.

The structural analysis of lexical bundles is useful because it shows how learners' academic writing develops from simple patterns toward more complex academic forms. Biber et al. (2021) classified lexical bundles into different structural types, including noun phrase bundles, prepositional phrase bundles, and verb-based bundles. In

academic writing, noun-based and preposition-based bundles are often linked with more formal and information-dense expression, such as the role of the, the purpose of this, and in terms of the. Recent studies suggest that advanced academic writing tends to contain more noun-based and prepositional structures because these forms allow writers to compress information and develop more formal arguments (Biber & Gray, 2016; Biber et al., 2021). In contrast, less proficient learners may rely more on verb-based or stance-based expressions, which can make their writing appear less academic. Functional classification is equally important because lexical bundles do not only have grammatical form; they also perform specific roles in discourse. Hyland (2008) classified lexical bundles into research-oriented, text-oriented, and participant-oriented categories. Research-oriented bundles help writers discuss content, procedures, evidence, and findings. Text-oriented bundles organize arguments and show logical relations between ideas. Participant-oriented bundles express stance or engage the reader. This framework is useful for examining ESL writing because it shows whether learners are using bundles only as memorized phrases or as meaningful academic resources. Recent work has also emphasized that the same bundle may function differently depending on its context, so frequency counts alone cannot fully explain learner development (Durrant, 2021; Hyland & Jiang, 2018).

The CEFR provides a useful framework for examining lexical bundle development because it allows learner writing to be compared across recognized proficiency levels. At B1 level, learners are generally expected to produce simple connected texts, so their lexical bundle use may remain limited, repetitive, and dependent on familiar connectors. At B2 level, learners are expected to explain viewpoints and organize arguments more clearly, which may lead to greater use of text-organizing bundles. At C1 level, learners are expected to write with greater fluency, precision, and control, so their use of lexical bundles may become more formal, varied, and academically suitable. The Council of Europe (2020) explains this development in terms of increasing control over language, organization, and communicative purpose. When lexical bundles are studied across B1, B2, and C1 levels, they can reveal how learners move from basic formulaic expression toward more controlled academic phraseology.

In the Pakistani ESL context, this area still needs more attention. English has a strong role in Pakistani higher education, but many learners continue to write under the influence of examination-based learning, memorized essays, limited exposure to authentic academic texts, and teacher-centered classroom practices. Recent Pakistani studies have pointed out that learners often struggle with academic vocabulary, writing organization, and effective use of English in formal contexts (Muhammad et al., 2023; Taimoor et al., 2021). Because of this, Pakistani ESL learners may use lexical bundles differently from learners in Western or East Asian contexts. The present article builds on this gap by examining lexical bundle development in Pakistani ESL learners' academic writing across B1, B2, and C1 levels, as outlined in the uploaded thesis.

Overall, the literature shows that lexical bundles are closely connected with academic writing proficiency. They support coherence, fluency, stance, and textual organization, but their use depends on proficiency level, genre, exposure, and instructional context. For Pakistani ESL learners, a CEFR-based corpus study can show how academic phraseology develops across levels and where learners need more support. This also has practical value for English language teaching in Pakistan because students need to learn not only individual words and grammar rules, but also the recurring phrase patterns through which academic writing is actually shaped.

Methodology

This study used a corpus-based research design to examine the development of lexical bundles in Pakistani ESL learners' academic writing across B1, B2, and C1 levels of the Common European Framework of Reference for Languages. A corpus-based design was considered suitable because lexical bundles are best examined through repeated patterns in actual learner writing rather than through intuition or isolated examples. Corpus methods allow the researcher to identify frequent word sequences, compare their distribution across proficiency levels, and interpret their structural and functional roles in academic discourse. This approach is consistent with earlier lexical bundle research, where frequency, recurrence, and distribution across texts are treated as important criteria for identifying meaningful phraseological patterns (Biber et al., 2021; Hyland, 2008). Recent studies have also shown that corpus-based analysis is useful for tracing how learner writing develops in terms of academic phraseology, coherence, and rhetorical control (Hyland & Jiang, 2018; Durrant, 2021).

The data for the study consisted of argumentative essays written by Pakistani ESL learners. Argumentative writing was selected because it is a common academic genre and requires learners to organize claims, support ideas with reasons, manage contrast, and present a clear position. These demands make it a suitable genre for studying lexical bundles because learners need recurring phraseological patterns to connect ideas and develop arguments. The study focused on learners placed at B1, B2, and C1 levels, as these levels represent movement from intermediate writing ability to more advanced academic expression. The CEFR was used as the proficiency framework because it provides a recognized basis for comparing learner performance across levels, particularly in terms of language control, coherence, and communicative purpose (Council of Europe, 2020). The uploaded thesis also follows this same direction by examining lexical bundles and keyness in Pakistani ESL learners' academic writing across B1, B2, and C1 levels.

The learner corpus was divided into three sub-corpora according to proficiency level: B1, B2, and C1. Each sub-corpus contained argumentative essays produced by learners belonging to that level. The purpose of this division was to make it possible to compare lexical bundle use across levels and observe whether learners' phraseological choices became more varied, formal, and academically controlled as proficiency increased. Since corpus size may differ from one level to another, all frequencies were normalized so that the results could be compared fairly. Normalization is important in corpus-based studies because raw frequency alone can be misleading when sub-corpora are not exactly equal in size. Therefore, bundle frequencies were calculated per standard word count to make the comparison more reliable. The composition of the learner corpus is presented in Table 1.

Table 1: Corpus Composition by CEFR Level

CEFR Level	Number of Essays	Approximate Word Count	Text Type	Learner Category	
B1	100	180,000 words	Argumentative essays	Intermediate Pakistani learners	ESL
B2	100	200,000 words	Argumentative essays	Upper-intermediate Pakistani learners	ESL

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

C1	100	190,000 words	Argumentative essays	Advanced Pakistani ESL learners
Total	300	570,000 words	Argumentative essays	Pakistani ESL learner corpus

Note. Frequencies were normalized per 100,000 words to make comparison across sub-corpora more reliable. The thesis reports B1 at 180,000 words, B2 at 200,000 words, and C1 at 190,000 words.

The study focused on three-word and four-word lexical bundles because these are commonly examined in academic writing research and provide a manageable yet meaningful unit of analysis. Three-word bundles often show basic recurring patterns, while four-word bundles can reveal more developed phraseological structures. The extraction of lexical bundles was carried out through corpus analysis software such as WordSmith Tools, which helps identify repeated sequences across the corpus. To avoid treating accidental repetitions as lexical bundles, the study applied frequency and dispersion criteria. Frequency helped identify repeated word sequences, while dispersion ensured that a bundle was not limited to only one or two essays but appeared across a wider range of learner texts. This follows the logic of corpus-based lexical bundle studies, where both frequency and distribution are used to separate meaningful recurrent patterns from isolated repetition (Biber et al., 1999; Hyland & Jiang, 2018). After extraction, the lexical bundles were manually checked to remove incomplete, irrelevant, or contextually unsuitable items. This step was necessary because automatic extraction tools can identify repeated strings, but they cannot always judge whether a sequence has meaningful discourse value. For example, a repeated phrase may occur because of the essay prompt rather than because of the learner's natural phraseological choice. Manual validation therefore helped improve the accuracy of the analysis. Recent corpus researchers have also stressed that automatic frequency results need careful interpretation because corpus tools can show patterns, but the researcher must still examine their context and function (Durrant, 2018; Biber et al., 2021).

The extracted lexical bundles were analyzed structurally and functionally. For structural analysis, the study followed the broad categories developed in earlier lexical bundle research, including noun phrase bundles, verb phrase bundles, prepositional phrase bundles, and other common grammatical patterns (Biber et al., 1999). This classification helped show whether learners at higher proficiency levels used more academic and information-dense structures. For example, noun-based and preposition-based bundles are often associated with formal academic writing, while lower-level writing may depend more on simple verb-based or stance-based expressions. Recent research has also linked academic writing development with a movement toward more compressed and noun-based structures, especially in advanced writing (Biber & Gray, 2016; Biber et al., 2021).

For functional analysis, the study used Hyland's (2008) taxonomy of lexical bundles, which classifies bundles into research-oriented, text-oriented, and participant-oriented functions. Research-oriented bundles refer to content, findings, procedures, or evidence. Text-oriented bundles help organize the argument, mark transitions, and show logical links between ideas. Participant-oriented bundles express the writer's stance or engage the reader. This classification was useful because lexical bundle development is not only about how often learners use bundles, but also about how effectively they use them in academic discourse. A learner may use many bundles, but if those bundles are

repetitive or functionally weak, they may not improve the quality of writing. Functional analysis therefore helped explain the rhetorical value of bundle use across B1, B2, and C1 levels.

The analysis was carried out in stages. First, the most frequent three-word and four-word lexical bundles were identified in each CEFR sub-corpus. Second, the frequency of these bundles was compared across B1, B2, and C1 levels to observe developmental movement. Third, the bundles were classified structurally to examine whether learners moved from simple and repetitive patterns toward more academic grammatical forms. Fourth, the bundles were classified functionally to see whether learners developed stronger control over text organization, evidence presentation, and stance expression. Finally, selected examples were examined in context to interpret how learners actually used these bundles in argumentative writing. This combination of quantitative frequency analysis and qualitative interpretation made the analysis more complete because it considered both numerical patterns and discourse function.

To maintain reliability, the study used clear criteria for bundle extraction, classification, and interpretation. The same frequency and dispersion principles were applied across all sub-corpora. Structural and functional categories were based on established frameworks from previous research, particularly Biber et al. (1999) and Hyland (2008). Manual checking was also used to reduce errors caused by automatic extraction. The purpose was to ensure that the findings reflected actual learner writing patterns rather than software noise or isolated textual repetition. Through this procedure, the study aimed to present a systematic account of how lexical bundles develop in Pakistani ESL learners' academic writing across CEFR levels.

Results and Discussion

The analysis of three-word and four-word lexical bundles showed a clear developmental movement across B1, B2, and C1 levels. The most frequent lexical bundles across the three CEFR levels are presented in Table 2.

Table 2: Most Frequent Lexical Bundles across CEFR Levels

CEFR Level	Lexical Bundle	Raw Frequency	Per 100,000 Words	Dispersion
B1	on the other hand	25	13.89	18%
B1	in my opinion	20	11.11	15%
B1	and also the	18	10.00	16%
B1	it is very	15	8.33	12%
B1	for example the	14	7.78	14%
B2	as a result	18	9.00	17%
B2	for example the	16	8.00	16%
B2	it is important	14	7.00	15%
B2	on the other hand	13	6.50	14%
B2	in this way	12	6.00	13%
C1	it is argued that	12	6.32	16%
C1	in terms of	10	5.26	15%

C1	the evidence suggests	9	4.74	13%
C1	it can be	8	4.21	12%
C1	as well as	7	3.68	11%

Note. B1 learners relied more on familiar and stance-based bundles, B2 learners showed more transitional and explanatory bundles, while C1 learners used more formal and academic phraseological patterns.

At the B1 level, learners relied heavily on familiar and easily available bundles such as in my opinion, on the other hand, and it is very. These bundles helped learners perform basic discourse functions, especially expressing personal stance and linking simple ideas, but their repeated use also showed a limited phraseological range. This pattern is expected at the intermediate level because learners are still developing control over academic expression and often depend on memorized classroom phrases. Biber et al. (1999) argued that lexical bundles act as repeated building blocks of discourse, but the present findings show that at lower proficiency levels these building blocks remain narrow and sometimes mechanically used. Similar patterns have been observed in learner writing, where less proficient writers often depend on a restricted set of high-frequency bundles rather than a wider academic repertoire (Chen & Baker, 2016; Hyland & Jiang, 2018).

At the B2 level, the learners showed a transitional pattern. Bundles such as as a result and for example became more noticeable, indicating that learners were beginning to organize ideas more clearly and support their arguments through examples and cause-effect relations. This level did not yet show full academic control, but it reflected movement away from purely personal or conversational phraseology. The presence of more text-organizing bundles suggests that B2 learners were developing a stronger sense of argument structure. This finding supports the view that lexical bundle development is not sudden; it takes place gradually as learners gain more exposure to academic texts and begin to use recurring phrases for more specific rhetorical purposes (Hyland & Jiang, 2018; Durrant, 2021). In this sense, B2 can be seen as a middle stage where learners move from basic sentence connection toward more organized academic writing.

At the C1 level, the bundle use became more formal, controlled, and academically oriented. Learners used bundles such as it is argued that and the evidence suggests, which show a movement toward impersonal stance, evidence-based reasoning, and more formal academic framing. These bundles are different from the B1 bundles because they do not merely express personal opinion or simple contrast. Instead, they help the writer present claims in a more cautious and academically acceptable way. This finding agrees with Hyland's (2008) view that academic writing depends on phraseological patterns that help writers organize claims, present evidence, and position themselves within discourse. It also supports recent work showing that advanced academic writing is marked by greater use of formulaic expressions that serve textual and research-based functions (Biber & Gray, 2016; Biber et al., 2021).

The structural analysis also showed development across proficiency levels. The structural classification of lexical bundles across B1, B2, and C1 levels is shown in Table 3.

Table 3: Structural Classification of Lexical Bundles across CEFR Levels

Bundle Type	B1	B2	C1	Developmental Pattern
Noun-based bundles	20	25	35	Gradual increase from B1 to C1
Verb-based bundles	35	30	30	Slight decline after B1
Preposition-based bundles	45	45	35	High at B1 and B2, lower at C1
Total	100	100	100	—

Note. The structural distribution shows a shift toward more noun-based academic forms at C1, while B1 and B2 learners depend more on verb-based and preposition-based structures. The chi-square result reported in the thesis was $\chi^2(4, N = 300) = 15.62, p < .01$, showing statistically significant differences across CEFR levels.

B1 learners used more simple and accessible structures, often built around stance expressions and basic verb-based patterns. Such bundles helped them continue sentences and connect ideas, but they did not always contribute to formal academic style. At B2 level, learners began to use a wider variety of structures, including prepositional and clause-based bundles. At C1 level, there was greater use of structurally denser patterns, especially noun-based and preposition-based bundles such as the role of and in terms of. This shift is important because academic writing often depends on compressed structures that allow writers to package information more formally. Biber and Gray (2019) noted that academic prose tends to rely heavily on noun phrase structures and compressed expression, and the present findings suggest that Pakistani ESL learners gradually move toward this pattern as their proficiency increases.

The functional analysis further confirmed that lexical bundle development is not only a matter of frequency. The functional classification of lexical bundles across CEFR levels is presented in Table 4.

Table 4: Functional Classification of Lexical Bundles across CEFR Levels

Function Type	B1	B2	C1	Notable Examples
Text-oriented bundles	30%	35%	40%	on the other hand; as a result; in terms of
Research-oriented bundles	20%	30%	35%	the evidence suggests; it is argued; based on the results
Participant-oriented bundles	50%	35%	25%	in my opinion; it is important; it is said
Total	100%	100%	100%	—

Note. The functional pattern shows that learners move from participant-oriented bundles at B1 toward more text-oriented and research-oriented bundles at C1. This indicates growing academic control and reduced dependence on direct personal stance. At B1 level, participant-oriented bundles were more common because learners often expressed personal views directly through phrases like in my opinion and I think that. These bundles are not wrong, but their frequent use can make academic writing appear subjective and less developed. At B2 level, text-oriented bundles became more visible, helping learners organize ideas through contrast, examples, and cause-effect relations. At C1 level, research-oriented and text-oriented bundles appeared more strongly, showing that learners were beginning to present arguments through evidence, interpretation, and academic distance. This movement from participant-oriented to

text-oriented and research-oriented bundles reflects increasing rhetorical control, which Hyland (2008) identifies as a major feature of academic phraseology. The comparison with BAWE also revealed important patterns of overuse and underuse. The overused and underused lexical bundles in comparison with BAWE are presented in Table 5.

Table 5: Overused and Underused Lexical Bundles Compared with BAWE

Lexical Bundle	CEFR Level	Type	Log-Likelihood Value	Interpretation
in my opinion	B1	Overrepresented	35.89	Strong reliance on personal stance
on the other hand	B1	Overrepresented	22.34	Heavy use of familiar contrast marker
it is very	B1	Overrepresented	19.11	General evaluative expression, less academically precise
the evidence suggests	C1	Underrepresented	-8.67	Limited use of evidence-based academic framing
it is argued that	C1	Underrepresented	-9.45	Limited use of impersonal academic claim-making
based on the results	C1	Underrepresented	-7.98	Weak use of research-oriented reporting bundle

Note. Positive LL values indicate overrepresentation in the Pakistani ESL learner corpus compared with BAWE, while negative LL values indicate underrepresentation. The thesis used WordSmith keyness analysis and Log-Likelihood statistics for this comparison.

The bundles *in my opinion*, *on the other hand*, and *it is very* were overrepresented in Pakistani ESL learner writing, especially at B1 level. The uploaded thesis reports Log-Likelihood scores of 35.89 for *in my opinion*, 22.34 for *on the other hand*, and 19.11 for *it is very*, showing that these bundles were strongly overused in comparison with BAWE academic writing. This suggests that Pakistani ESL learners, particularly at lower proficiency levels, often depend on familiar classroom phrases to organize their writing. While these expressions may help learners produce connected texts, excessive use can reduce variety and weaken the academic tone of writing.

At the same time, some more academic bundles were underrepresented when compared with BAWE. Bundles such as *the evidence suggests*, *it is argued that*, and *based on the results* appeared less frequently than expected in advanced academic writing. The thesis identifies these as underrepresented C1 bundles, with negative Log-Likelihood values for *the evidence suggests*, *it is argued that*, and *based on the results*. This finding is important because it shows that even advanced Pakistani ESL learners may not fully match proficient academic writing patterns. Their writing improves across CEFR levels,

but certain evidence-based and research-oriented expressions still need more classroom attention. This supports the argument that academic phraseology should be taught explicitly rather than left to incidental learning (Hyland & Jiang, 2018; Durrant, 2021). Overall, the findings show that lexical bundle development among Pakistani ESL learners moves from personal, repetitive, and conversational patterns at B1 level toward more organized and academically shaped patterns at B2 and C1 levels. The thesis summary also confirms this movement, noting that B1 learners frequently used stance markers and basic connectors, B2 learners showed greater use of transitional bundles, and C1 learners moved toward academic register through bundles such as it is argued that and the evidence suggests. This progression indicates that lexical bundles are useful indicators of academic writing development. They reveal not only what learners write frequently, but also how they organize argument, express stance, and build academic authority. This developmental movement across CEFR levels is summarized in Figure 1.

CEFR Level	Dominant Bundle Pattern	Main Writing Function	Developmental Meaning
B1	Personal and repetitive bundles: in my opinion, on the other hand, it is very	Expressing basic stance and simple connection	Learners depend on familiar classroom phrases and show limited academic phraseology
↓			
B2	Transitional and text-organizing bundles: as a result, for example the, in this way	Explaining, exemplifying, and linking arguments	Learners begin to organize ideas more clearly and move toward academic discourse
↓			
C1	Formal and academic bundles: it is argued that, the evidence suggests, in terms of	Presenting claims, evidence, and impersonal academic stance	Learners show stronger rhetorical control and more mature academic phraseology

Figure 1: Developmental Movement of Lexical Bundles across CEFR Levels

The figure shows a gradual shift from personal and repetitive bundle use at B1, to transitional and text-organizing bundles at B2, and finally to more formal, evidence-based, and research-oriented bundles at C1.

These findings are consistent with previous studies which argue that lexical bundle use reflects proficiency, rhetorical control, and familiarity with academic conventions. Staples et al. (2013) showed that learner development can be traced through changes in frequency, function, and distribution of formulaic sequences. Chen and Baker (2016) also found that learner writers differ from more proficient academic writers in their structural and functional use of bundles. The present study adds to this discussion by providing evidence from Pakistani ESL learners, a context that has received limited attention in corpus-based lexical bundle research. It shows that Pakistani learners do develop stronger academic phraseology across CEFR levels, but the development remains uneven because some familiar bundles are overused while more academic bundles remain underused.

The results therefore suggest that lexical bundle development should be treated as an important part of ESL academic writing instruction. Learners do not only need

grammar correction or vocabulary expansion; they also need guided exposure to the recurring phrase patterns used in academic discourse. At B1 level, learners may need support in moving beyond personal expressions such as *in my opinion*. At B2 level, instruction can focus on using text-organizing bundles more accurately and flexibly. At C1 level, learners need more practice with research-oriented and evidence-based bundles so that their writing becomes closer to formal academic norms. This is especially relevant in Pakistan, where learners often write English under examination pressure and may not receive enough exposure to authentic academic texts (Muhammad et al., 2023; Taimoor et al., 2021).

Pedagogical Implications

The findings of this study suggest that lexical bundles should be taught more directly in Pakistani ESL writing classrooms. Many learners appear to know some common bundles, but their use is often narrow, repetitive, and limited to basic connectors or personal stance expressions. This means that learners may be able to write connected sentences, yet they may still struggle to produce academic arguments with enough fluency and formal control. Hyland and Jiang (2018) argue that academic writing depends strongly on recurring phraseological patterns, and this point is useful for Pakistani classrooms because students often learn grammar and vocabulary separately but receive less training in how academic phrases work inside real texts. If lexical bundles are taught as part of writing instruction, learners can begin to see how academic writers introduce claims, build contrast, present evidence, and link ideas.

At the B1 level, teaching should focus on helping learners move beyond a small set of personal and overused bundles such as *in my opinion*, *I think that*, and *on the other hand*. These expressions are useful at an early stage, but if learners continue to depend on them too heavily, their writing may sound informal and repetitive. Teachers can introduce simple text-oriented bundles such as *for example*, *as a result*, *in this way*, and *in the case of* so that learners can organize ideas more clearly. At this level, instruction should not overload students with complex academic phrases. It should begin with controlled practice, short writing tasks, and sentence-completion activities where students learn how one bundle performs one clear function in a paragraph. This matches the CEFR view that B1 learners are still developing the ability to write simple connected texts with reasonable control (Council of Europe, 2020).

At the B2 level, learners need support in using bundles for argument development rather than only sentence connection. Since B2 learners are expected to explain viewpoints and present more detailed writing, they should be trained to use bundles that help them compare, explain, justify, and qualify ideas. Expressions such as *as a result of*, *in relation to*, *it can be seen*, *this shows that*, and *one of the main* can help learners build more organized arguments. However, teachers should also show students that lexical bundles must fit the meaning of the sentence. For example, *on the other hand* should only be used when there is a real contrast, not simply as a general paragraph opener. Durrant (2021) notes that learners benefit when phraseological instruction is connected with context, because memorized phrases alone do not guarantee better academic writing.

At the C1 level, instruction should shift toward more formal, research-oriented, and evidence-based bundles. Learners at this stage need practice with expressions such as *the findings suggest that*, *it is argued that*, *the evidence indicates*, *based on the results*, and *the purpose of this*. These bundles help writers present claims with academic distance and support arguments through evidence rather than direct personal opinion.

Biber and Gray (2019) point out that advanced academic writing often relies on compressed and formal structures, especially noun-based and prepositional patterns. Therefore, C1 learners should be guided to use bundles that make their writing less conversational and more aligned with academic discourse. This is especially important for Pakistani ESL learners who may reach higher proficiency but still retain classroom-based or examination-based phrase patterns.

The study also suggests that textbook writers and curriculum designers in Pakistan should give more space to academic phraseology. Current writing materials often give lists of connectors or model essays, but they may not explain how recurring phrases function in actual academic writing. Lexical bundles can be introduced in writing units according to function: bundles for introducing arguments, bundles for contrast, bundles for cause and effect, bundles for evidence, bundles for cautious claims, and bundles for concluding a point. Such organization would help learners understand that bundles are not decorative expressions; they are tools for meaning-making. Hyland (2008) explains that lexical bundles support research-oriented, text-oriented, and participant-oriented functions, and this classification can be adapted for classroom use in a simple way.

Corpus-informed teaching can also improve lexical bundle instruction. Teachers can use short concordance examples from learner writing and expert academic writing to show how bundles are used in context. For instance, students can compare repeated learner use of in my opinion with more academic alternatives such as it can be argued that or the evidence suggests that. This kind of comparison can make students more aware of overuse, underuse, and appropriateness. Chen and Baker (2016) showed that learner writing often differs from proficient academic writing in both range and function of lexical bundles, so classroom comparison between learner and expert patterns can help students notice these differences. Even if full corpus tools are not available in every classroom, teachers can still use small sets of example sentences taken from academic essays, research articles, or student assignments.

Teacher training is also needed because lexical bundles may be unfamiliar to many instructors who were trained through traditional grammar-based methods. Teachers should be introduced to the idea that writing proficiency is not built only through correct grammar and individual vocabulary, but also through recurring patterns of academic expression. Training workshops can show teachers how to identify frequent bundles, explain their functions, and design writing tasks around them. In Pakistani ESL classrooms, this can be done practically through paragraph rewriting, bundle substitution, peer review, and comparison of weak and strong academic paragraphs. Such activities can help students understand why one phrase works better than another in a given academic context.

The findings also have implications for assessment. If lexical bundle use reflects writing development across CEFR levels, then assessment rubrics should pay some attention to phraseological control. This does not mean that students should be rewarded for inserting many memorized phrases. Rather, assessors should look at whether learners use recurring academic expressions appropriately, flexibly, and in relation to the argument. A B1 learner may be assessed for basic connection of ideas, a B2 learner for clearer organization and explanation, and a C1 learner for formal academic stance and evidence-based expression. This would make writing assessment more sensitive to actual developmental differences in ESL academic writing.

Overall, the pedagogical value of this study lies in showing that lexical bundles can help Pakistani ESL learners move from basic connected writing toward more formal

and academically controlled expression. Instead of treating bundles as fixed phrases to be memorized, teachers should present them as functional resources that shape argument, coherence, and academic voice. This approach can make writing instruction more practical and more closely linked with the language patterns students need in higher education.

Conclusion

This study examined lexical bundle development in Pakistani ESL learners' academic writing across B1, B2, and C1 CEFR levels through a corpus-based approach. The findings show that lexical bundles are closely linked with writing proficiency because they help learners organize ideas, express stance, connect arguments, and gradually move toward more formal academic expression. B1 learners mostly relied on simple and repetitive bundles, particularly personal stance markers and basic connectors, while B2 learners showed a transitional pattern through more text-organizing bundles. At the C1 level, learners demonstrated stronger control over academic bundles related to evidence, interpretation, and impersonal stance, showing a clear movement from basic phrase use toward more mature academic phraseology. The study concludes that lexical bundle development is not only a matter of frequency, but also of structure, function, and contextual appropriateness. Pakistani ESL learners need explicit instruction in academic phraseology so that they can move beyond memorized expressions and use lexical bundles more effectively in argument building. This has direct relevance for ESL writing classrooms in Pakistan, where teachers can use model texts, corpus-based examples, and paragraph-level practice to improve learners' coherence, fluency, and academic control. Future studies may extend this work by examining larger corpora, other genres, discipline-specific writing, or longitudinal development of lexical bundles over time.

References

- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of spoken and written English*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.232>
- Chen, Y.-H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849–880. <https://doi.org/10.1093/applin/amu065>
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment: Companion volume*. Council of Europe Publishing.
- Durrant, P. (2018). Formulaic language in English for academic purposes. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 211–227). Routledge. <https://doi.org/10.4324/9781315206615-12>
- Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding development and proficiency in writing: Quantitative corpus linguistic approaches*. Cambridge University Press. <https://doi.org/10.1017/9781108770101>

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

- Gurmani, M. T. A. A., Yusoff, Z. S., & Shahid, C. (2021). An evaluation of vocabulary of English textbooks of Pakistan and Saudi Arabia through CEFR. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(10), 3375–3387.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Jiang, F. (2019). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383–407. <https://doi.org/10.1075/ijcl.17080.hyl>
- Muhammad, R., Shahid, C., & Gurmani, T. A. (2023). Analysis of the factors affecting the implementation of communicative language teaching approach in Pakistani private and public institutions. *Journal of Languages, Culture and Civilization*, 5(1), 17–29. <https://doi.org/10.47067/jlcc.v5i1.164>
- Nesi, H., Gardner, S., Thompson, P., & Wickens, P. (2004). *British Academic Written English Corpus*. Oxford Text Archive.
- Scott, M. (2021). *WordSmith Tools* (Version 8) [Computer software]. Lexical Analysis Software.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214–225. <https://doi.org/10.1016/j.jeap.2013.05.002>